

SKRIPSI

**IMPLEMENTASI ALGORITMA *NAIVE BAYES CLASSIFIER*
UNTUK MENGANALISIS KEPERIBADIAN PENGGUNA
TWITTER DI INDONESIA BERDASARKAN PENDAPATNYA
TERHADAP PASANGAN CALON PRESIDEN DAN WAKIL
PRESIDEN TAHUN 2019**



TESYA NURINTAN

15650022

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2019

**IMPLEMENTASI ALGORITMA *NAIVE BAYES CLASSIFIER*
UNTUK MENGANALISIS KEPERIBADIAN PENGGUNA
TWITTER DI INDONESIA BERDASARKAN PENDAPATNYA
TERHADAP PASANGAN CALON PRESIDEN DAN WAKIL
PRESIDEN TAHUN 2019**

SKRIPSI

Untuk memenuhi sebagian persyaratan
mencapai derajat Sarjana S-1
Program Studi Teknik Informatika



Disusun Oleh

TESYA NURINTAN

15650022

Kepada

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
YOGYAKARTA**

2019



KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN KALIJAGA
FAKULTAS SAINS DAN TEKNOLOGI

Jl. Marsda Adisucipto Telp. (0274) 540971 Fax. (0274) 519739 Yogyakarta 55281

PENGESAHAN TUGAS AKHIR

Nomor : B-2028/Un.02/DST/PP.00.9/06/2019

Tugas Akhir dengan judul : IMPLEMENTASI ALGORITMA NAIVE BAYES CLASSIFIER UNTUK
MENGANALISIS KEPERIBADIAN PENGGUNA TWITTER DI INDONESIA
BERDASARKAN PENDAPATNYA TERHADAP PASANGAN CALON PRESIDEN
DAN WAKIL PRESIDEN TAHUN 2019

yang dipersiapkan dan disusun oleh:

Nama : TESSYA NURINTAN
Nomor Induk Mahasiswa : 15650022
Telah ditujikan pada : Jumat, 24 Mei 2019
Nilai ujian Tugas Akhir : A-

dinyatakan telah diterima oleh Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta

TIM UJIAN TUGAS AKHIR

Ketua Sidang

Muhammad Didik Rahmad Wahyudi, S.T., MT.
NIP. 19760812 200901 1 015

Penguji I

Dr. Agung Fatwanto, S.Si., M.Kom.
NIP. 19770103 200501 1 003

Penguji II

Agus Mulyanto, S.Si., M.Kom.
NIP. 19710823 199903 1 003

Yogyakarta, 24 Mei 2019
UIN Sunan Kalijaga
Fakultas Sains dan Teknologi
B E K A N



Dr. Murtoto, M.Si.
NIP. 19691212 200003 1 001

SURAT PERSETUJUAN SKRIPSI/TUGAS AKHIR

Hal :
Lamp :

Kepada
Yth. Dekan Fakultas Sains dan Teknologi
UIN Sunan Kalijaga Yogyakarta
di Yogyakarta

Assalamu'alaikum wr. wb.

Setelah membaca, meneliti, memberikan petunjuk dan mengoreksi serta mengadakan perbaikan seperlunya, maka kami selaku pembimbing berpendapat bahwa skripsi Saudara:

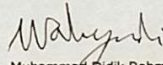
Nama : Tesya Nurintan
NIM : 15650022
Judul Skripsi : Analisis Karakter Masyarakat Indonesia Terhadap Pasangan Calon Presiden dan Calon Wakil Presiden Tahun 2019 Berdasarkan Tweet Menggunakan Naive Bayes Classifier.
Algoritma

sudah dapat diajukan kembali kepada Program Studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu dalam Teknik Informatika

Dengan ini kami berharap agar skripsi/tugas akhir Saudara tersebut di atas dapat segera dimunaqsyahkan. Atas perhatiannya kami ucapkan terima kasih.

Wassalamu'alaikum wr. wb.

Yogyakarta, 24 Mei 2019
Pembimbing



Muhammad Didik Rohmad Wahyudi, S.T., MT.
NIP. 19760812 200901 1 015

PERNYATAAN KEASLIAN SKRIPSI

Saya yang bertanda tangan dibawah ini :

Nama : Tesya Nurintan

NIM : 15650022

Jurusan : Teknik Informatika

Fakultas : Sains dan Teknologi

Menyatakan bahwa skripsi saya yang berjudul "**Analisis Karakter Masyarakat Indonesia Terhadap Pasangan Calon Presiden dan Calon Wakil Presiden Tahun 2019 Berdasarkan *Tweet* Menggunakan Algoritma *Naïve Bayes Classifier*.**" merupakan hasil penelitian saya sendiri, tidak terdapat pada karya yang pernah di ajukan untuk memperoleh gelar kesarjana di suatu perguruan tinggi, dan bukan plagiasi karya orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 23 Mei 2019



Menyatakan,

Tesya Nurintan
NIM.15650022

KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan rahmat dan hidayah-Nya sehingga penulis bisa menyelesaikan penelitian yang berjudul **Implementasi Algoritma Naive Bayes Classifier Untuk Menganalisis 5 Besar Karakter Kepribadian Masyarakat Indonesia Terhadap Pasangan Calon Presiden dan Wakil Presiden Tahun 2019 Berdasarkan Data Tweets**. Sholawat dan salam selalu tercurahkan kepada junjungan umat muslim, Nabi Muhammad SAW. beserta keluarga, sahabat, dan kerabat beliau. Semoga kita semua mendapatkan *syafaat* di hari akhir kelak. Amin. Penulis menyadari bahwa terdapat banyak kendala dalam pelaksanaan dan penulisan penelitian ini. Namun berkat bantuan dan kerja sama dari berbagai pihak, kendala-kendala tersebut dapat diatasi. Untuk itu penulis mengucapkan terima kasih dan penghargaan kepada :

1. Bapak Prof. Drs. K.H. Yudian Wahyudi, M.A., Ph.D., selaku Rektor UIN Sunan Kalijaga Yogyakarta.
2. Bapak Dr. Murtono, M.Si., selaku Dekan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
3. Bapak Sumarsono, S.T., M.Kom., selaku Ketua Program Studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta.
4. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku dosen pembimbing skripsi yang telah sabar dan meluangkan waktunya untuk memberikan motivasi, koreksi dan kritik saran kepada penulis sehingga skripsi ini dapat diselesaikan.

5. Bapak Dr. Bambang Sugiantoro, M.T., selaku Dosen Pembimbing Akademik.
6. Dr. Agung Fatwanto, S.Si., M.Kom., Agus Mulyanto, S.Si., M.Kom., Ph.D., Aulia Faqih Rifa'i, M.Kom., M. Taufiq NUruzzaman, S.T., Maria Ulfah Si-regar, S.Kom. MIT., Ph.D., Nurochman, S.Kom., M.Kom., Rahmat Hidayat, S.Kom., M.Cs., Dr. Shofwatul 'Uyun, S.T., M.Kom., selaku dosen pengampu mata kuliah program studi Teknik Informatika UIN Sunan Kalijaga Yogyakarta yang telah banyak membantu penulis hingga penulis dapat menyusun tugas akhir.
7. Seluruh staf dan karyawan Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta.
8. Tim Drone Emprit yang telah memberikan akses kepada penulis untuk mengunduh dan mengolah data yang terdapat di sistem Drone Emprit.
9. Orang tua dan keluarga yang selalu memberikan dukungan tanpa pamrih kepada penulis.
10. Teman-teman program studi Teknik Informatika Angkatan 2015 UIN Sunan Kalijaga Yogyakarta
11. Semua pihak yang telah membantu dalam proses pengerjaan dan penulisan penelitian ini yang tidak bisa penulis sebutkan satu per satu.

Semoga Allah SWT membalas kebaikan dari seluruh pihak yang telah membantu penulis dalam pelaksanaan dan penyelesaian penelitian ini. Akhir kata, penulis menyadari bahwa penelitian ini bahwa masih terdapat banyak kesalahan dan

jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang dapat berguna dalam penyempurnaan penelitian ini di masa mendatang. Semoga penelitian ini dapat bermanfaat bagi pembaca.

Yogyakarta, 10 Mei 2019

Penulis

HALAMAN PERSEMBAHAN

Alhamdulillah *Alamin*, segala puji syukur penulis panjatkan kehadiran Allah SWT atas rahmat dan hidayah-Nya yang tak terhingga sehingga penulis dapat menyelesaikan penelitian yang berjudul **Implementasi Algoritma Naive Bayes Classifier Untuk Menganalisis 5 Besar Karakter Kepribadian Masyarakat Indonesia Terhadap Pasangan Calon Presiden dan Wakil Presiden Tahun 2019 Berdasarkan Data Tweets**. Sholawat dan salam senantiasa tercurahkan kepada kekasih Allah SWT, junjungan umat muslim sepanjang masa, sumber inspirasi penulis, Nabi Muhammad SAW. beserta keluarga, sahabat, dan kerabat yang menemani perjalanan hidup beliau. Semoga kelak kita semua mendapatkan syafaat dari beliau dan dapat berkumpul kembali di *Jannah*-Nya. Amin.

Penulis ingin mengucapkan terima kasih dan apresiasi kepada semua pihak yang telah memberikan dukungan dan motivasi kepada penulis sehingga penelitian ini dapat penulis selesaikan dengan baik, diantaranya :

1. Kedua orang tua tercinta yang selalu mendoakan serta memberikan dukungan moril dan materil sehingga penulis mampu melangkah menggapai cita-cita penulis hingga sejauh ini.
2. Kedua adik penulis, Muhammad Ihsan Hidayah dan Assyifatul Haifa Shaliha yang selalu menjadi penyemangat bagi penulis untuk terus berusaha meraih cita-cita.
3. Bapak Muhammad Didik Rohmad Wahyudi, S.T., MT., selaku dosen pembimbing skripsi yang telah banyak memberikan saran, motivasi, serta bimbingan.

bingan dalam proses pelaksanaan dan penulisan penelitian ini.

4. Segenap Bapak dan Ibu dosen Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga. Terima kasih atas semua ilmu dan pengalaman yang diberikan kepada penulis semasa kuliah.
5. Bapak Aulia Faqih Rifa'i, M.Kom., Tim Dirakit dan Paguyuban Lab selaku dosen pembimbing serta kakak senior yang telah membimbing dan menukarkan ilmu serta pengalaman kepada penulis sejak awal bergabung di *lab*.
6. Teman-teman *After Sunday*, *DSC SUKA Core Team*, *Developer Student Clubs (DSC)* Indonesia dan *Team Googler* Indonesia.
7. *Encouraging Friends* penulis : Ambar, Meva, dan Alfin.
8. Yuk Bama dan Kak Ari yang telah membantu penulis dalam penelitian ini.
9. Niken (*my undergrad-thesis partner*) serta temen Nemoku Lana, Andri, dan Umam sebagai teman lembur skripsian.
10. Semua pihak yang telah banyak membantu dalam penyelesaian skripsi yang tidak bisa penulis sebutkan satu persatu.

Semoga Allah SWT membalas kebaikan dari seluruh pihak yang telah membantu penulis dalam pelaksanaan dan penyelesaian penelitian ini.

"Laa Hawla wa Laa Quwwata Illa Billah."

"For when Allah SWT is our strength, nothing can break us. Nothing."

- - -

"... Surely, my Lord hears all prayers."

(QS. 14:39)

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN SKRIPSI/TUGAS AKHIR	ii
LEMBAR PERSETUJUAN SKRIPSI/TUGAS AKHIR	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
KATA PENGANTAR	v
HALAMAN PERSEMBAHAN	viii
HALAMAN MOTTO	x
DAFTAR ISI	xi
DAFTAR TABEL	xiv
DAFTAR TABEL	xv
DAFTAR LAMBANG	xvi
INTISARI	xvii
ABSTRACTxviii
I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah	3
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
1.6. Keaslian Penelitian	5
1.7. Sistematika Penulisan	5
II TINJAUAN PUSTAKA DAN LANDASAN TEORI	7

2.1. Tinjauan Pustaka	7
2.2. LANDASAN TEORI	13
2.2.1. Data Mining	13
2.2.2. Klasifikasi	15
2.2.3. <i>Supervised Learning</i>	15
2.2.4. Analisis sentimen	15
2.2.5. <i>Naive Bayes Classifier (NBC)</i>	16
2.2.6. TF dan TF-IDF	19
2.2.7. <i>K-Fold Cross Validation</i>	19
2.2.8. <i>Python</i>	20
2.2.9. Lima Besar Ciri Kepribadian (<i>Big 5 Personality</i>)	20
III METODE PENELITIAN	22
3.1. Metode Penelitian	22
3.2. Tahap-tahap Penelitian	22
3.3. Studi Pustaka	23
3.4. Pengumpulan Data	23
3.5. Seleksi Data (<i>Data Selection</i>)	23
3.6. <i>Preprocessing</i>	23
3.7. Pembobotan Kata	24
3.8. <i>Model Classifier</i>	24
3.9. Analisis dan Evaluasi	24
IV HASIL DAN PEMBAHASAN	25
4.1. Pengumpulan Data	26

4.2. Seleksi Data	29
4.2.1. Menghilangkan Url	29
4.2.2. Menghapus Duplikasi	30
4.3. <i>Labeling</i>	31
4.4. <i>Preprocessing</i>	33
4.4.1. <i>Cleansing</i>	33
4.4.2. <i>Casefolding</i>	35
4.4.3. <i>Tokenizing</i>	37
4.4.4. <i>Convert Slang Word</i>	39
4.4.5. <i>Convert Negation</i>	41
4.4.6. <i>Stopword Removal</i>	43
4.4.7. <i>Stemming</i>	45
4.5. Analisis dan Evaluasi	47
4.5.1. Analisis	47
4.5.2. <i>Term Frequency - Inverse Document Frequency (TF-IDF)</i>	57
4.5.3. Evaluasi Model Klasifikasi	59
4.5.4. Implementasi	65
V PENUTUP	70
5.1. Kesimpulan	70
5.2. Saran	70
DAFTAR PUSTAKA	72
A CONTOH DATA PENELITIAN	74
CURICULUM VITAE	80

DAFTAR TABEL

2.1	Penelitian-penelitian Sebelumnya	10
4.1	Kata Kunci Data <i>Tweets</i>	27
4.2	Contoh Data <i>tweets</i>	27
4.4	Contoh Data Yang Mengandung Url	29
4.5	Contoh Data Hasil Penghapusan Url	30
4.7	Contoh <i>Labeling</i> Data <i>tweets</i>	32
4.8	Contoh <i>Cleansing</i> Data	33
4.9	Contoh <i>Casefolding</i> Data	35
4.10	Contoh <i>Tokenizing</i> Data	37
4.11	Contoh Data Kamus <i>Slang Word</i>	39
4.12	Contoh <i>Convert Slang Word</i>	40
4.13	Contoh Data Hasil Proses <i>Convert Negation</i>	42
4.14	Contoh Data Hasil Proses <i>Stopword Removal</i>	43
4.15	Contoh Data Hasil Proses <i>Stemming</i>	45
4.16	Contoh Data Hasil Proses Seleksi, <i>Labeling</i> , dan <i>Preprocessing</i>	48
4.17	<i>Term Frequency</i>	49
4.18	Peluang <i>Term</i> Kelas <i>Openness</i>	51
4.19	Peluang <i>Term</i> Kelas <i>Conscientiousness</i>	51
4.20	Peluang <i>Term</i> Kelas <i>Extraversion</i>	52
4.21	Peluang <i>Term</i> Kelas <i>Agreeableness</i>	53
4.22	Peluang <i>Term</i> Kelas <i>Neuroticism</i>	54
4.23	<i>Perhitungan TF-IDF</i>	58

DAFTAR GAMBAR

3.1 Tahapan Penelitian	22
4.1 Proses Analisis	26
4.2 Contoh Data Duplikat	31
4.3 Grafik Akurasi K-fold cross validation TF	60
4.4 Grafik Akurasi K-fold cross validation TF-IDF	60
4.5 Hasil Klasifikasi	66
1.1 Contoh Data Mentah Sebelum dilakukan Seleksi Data	74
1.2 Contoh Data Mentah Setelah dilakukan Penghapusan Url	75
1.3 Contoh Data Mentah yang masih terdapat duplikasi data	76
1.4 Contoh Data Mentah yang sudah dilakukan sortir <i>unique</i>	77
1.5 Contoh Data Setelah Diseleksi dan Diberi Label	78
1.6 Contoh Data Setelah Dilakukan <i>Preprocessing</i>	79

DAFTAR LAMBANG

- $x \in A$: x anggota A
- $A \subseteq X$: A himpunan bagian (*subset*) atau sama dengan X
- \mathbb{N} : himpunan semua asli
- \mathbb{Z} : himpunan semua bilangan bulat
- \mathbb{Z}^+ : himpunan semua bilangan bulat positif
- \mathbb{R} : himpunan semua bilangan real
- C_r^n : r -kombinasi dari n unsur yang berbeda
- : akhir suatu bukti
- : akhir suatu contoh
- \rightarrow : menuju
- $\sum_{i=1}^n a_i$: penjumlahan $a_1 + a_2 + \cdots + a_n$
- $\prod_{i=1}^n a_i$: perkalian $a_1 \cdot a_2 \cdot \cdots \cdot a_n$
- $p \Rightarrow q$: jika p maka q
- \Leftrightarrow : jika dan hanya jika
- $x \leftarrow a$: nilai a dimasukkan ke x

INTISARI

IMPLEMENTASI ALGORITMA *NAIVE BAYES CLASSIFIER* UNTUK MENGANALISIS KEPRIBADIAN PENGGUNA TWITTER DI INDONESIA BERDASARKAN PENDAPATNYA TERHADAP PASANGAN CALON PRESIDEN DAN WAKIL PRESIDEN TAHUN 2019

Oleh

Tesya Nurintan

15650022

Etnografi merupakan sebuah studi yang dilakukan untuk mengetahui perilaku suatu masyarakat yang diobservasi langsung dari kelompok atau masyarakat yang diteliti itu sendiri. Twitter sebagai salah satu platform media sosial dapat dijadikan sumber untuk melakukan penelitian. Sebab masyarakat Indonesia tidak hanya memanfaatkan twitter untuk bersosialisasi, namun juga untuk melakukan kegiatan politik yakni menjadikannya tempat untuk mencurahkan pendapat mereka terhadap pasangan calon presiden dan calon wakil presiden.

Penelitian ini bermaksud untuk melakukan analisis terhadap personal karakter dari pengguna twitter di Indonesia mengenai pandangannya terhadap pasangan calon Presiden dan calon Wakil Presiden tahun 2019 dengan metode *Naive Bayes Classifier*. Personal karakter yang ingin diketahui dari penelitian ini yakni 5 Besar Kelompok Kepribadian, yaitu : *Openness*, *Conscientiousness*, *Extraversion*, *Agreeableness*, *Neuroticism*. Data didapatkan dari Sistem Drone Emprit dan diproses untuk dilakukan klasifikasi dengan tahapan seleksi data, labeling, preprocessing, pembobotan, dan klasifikasi.

Hasil yang diperoleh dari peneilitan ini adalah bahwa algoritma *Naive Bayes Classifier* dapat digunakan untuk mengklasifikasikan data *tweet* dengan tingkat akurasi yang lebih tinggi dengan pembobotan TF yaitu 39,96% dibandingkan dengan menggunakan metode pembobotan TF-IDF yakni 38,45%. Hasil implementasi algoritma *Naive Bayes Classifier* menghasilkan pengolompokkan kelas *Openness* sebanyak 42,4%, *Conscientiousness* sebanyak 9%, *Extraversion* sebanyak 14,9%, *Agreeableness* sebanyak 24%, dan *Neuroticism* sebanyak 9,7%.

Kata Kunci : *Naive Bayes*, Analisis 5 Besar Karakter Seseorang, Pemilihan Presiden dan Wakil Presiden, *k-fold cross validation*

ABSTRACT

THE IMPLEMENTATION OF NAIVE BAYES CLASSIFIER ALGORITHM FOR ANALYZING THE BIG 5 PERSONALITY OF INDONESIAN USER OF TWITTER TOWARDS THE PRESIDENT AND VICE-PRESIDENT CANDIDATES IN 2019 BASED ON THEIR OPINION

By

Tesya Nurintan

15650022

Ethnography is a study to determine the behavior of a community that is observed directly from the group or community studied itself. Twitter as one of the social media platforms can be used as a source for conducting research. Because of this technology era, people use twitter not only socializing, but also people use Twitter to do political activities, people use Twitter as a platform to express their opinions to the presidential candidates.

This research aims to analyze the personal character of the twitter users in Indonesia regarding their views on the President and Vice-President candidates in 2019 based on their opinion using the Naive Bayes Classifier algorithm. The personal character is the Big 5 Personality . They are: Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism. The data was obtained from Drone Emprit and processed to get a classification of Big 5 Personality with stages of data selection, labeling, preprocessing, weighting, and classification.

The results obtained from this study are that the Naive Bayes Classifier algorithm can be used to classify tweet data with a higher level of accuracy 39.96% with the TF method compared to 38.45% with the TF-IDF method. The results of implementing the Naive Bayes Classifier algorithm resulted in 42.4% Openness, 9% Conscientiousness, 14.9% Extraversion, 24% Agreeableness, and 9.7% Neuroticism.

Keywords : *Naive Bayes*, The Big 5 Personality, Presidential Election, k-fold cross validation.

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pemilihan Umum atau pemilu tahun 2019 diselenggarakan pada tanggal 17 April 2019. Tahun ini, masyarakat Indonesia mengikuti pemilu serentak Presiden dan Wakil Presiden, DPD RI, DPR RI, DPRD Provinsi, serta DPRD Kabupaten / Kota. Ajang pemilihan Presiden dan Wakil Presiden atau juga dikenal sebagai pesta demokrasi selalu menyita perhatian dari masyarakat Indonesia. Masyarakat Indonesia beramai-ramai mengutarakan ekspresi atau pendapat mereka dengan berbagai macam cara. Salah satunya dengan berkomentar melalui media sosial. Twitter sebagai salah satu platform media sosial yang ramai digunakan oleh masyarakat Indonesia, dapat dijadikan sumber untuk melakukan penelitian. Sebab, di era teknologi seperti sekarang, masyarakat menggunakan sosial media hampir di dalam setiap rangkaian kegiatan sehari-hari mereka. Tidak hanya membagikan atau menceritakan kegiatan mereka melalui postingan tweet dan bersosialisasi dengan orang lain, masyarakat juga memanfaatkan twitter untuk melakukan kegiatan politik yakni memanfaatkannya sebagai tempat untuk mencurahkan pendapat mereka terhadap pasangan calon presiden dan calon wakil presiden.

Untuk mengobservasi kepribadian seseorang, dapat dilakukan dengan banyak cara baik secara kualitatif maupun kuantitatif. (Haris, 2010) menjelaskan setidaknya ada 5 metodologi dalam penelitian kualitatif, salah satunya adalah etnografi. Etnografi merupakan sebuah studi yang dilakukan untuk mengetahui perilaku suatu

masyarakat yang diobservasi langsung dari kelompok atau masyarakat yang diteliti itu sendiri. Etnografi berasal dari kata *ethnies* dan *graph*. Kata-kata tersebut diartikan sebagai suatu studi atau penelitian yang difokuskan pada penjelasan deskriptif dan interpretasi terhadap budaya dan sistem sosial suatu kelompok atau suatu masyarakat tertentu melalui pengamatan dan penghayatan secara langsung terhadap suatu kelompok atau masyarakat yang diteliti. Peneliti memfokuskan penelitiannya pada kelompok atau suatu masyarakat tertentu yang dapat diobservasi dan pola pembelajaran suatu perilaku, adat istiadat, dan pandangan hidup kelompok tersebut melalui pengamatan dan penelitian secara langsung terhadap kelompok yang bersangkutan (Haris, 2010).

Penelitian ini bermaksud untuk melakukan analisis terhadap personal karakter dari masyarakat Indonesia mengenai pandangan mereka terhadap pasangan calon Presiden dan calon Wakil Presiden tahun 2019 berdasarkan data tweet mereka di platform media sosial twitter dengan menggunakan metode *Naive Bayes Classifier*. Personal karakter yang ingin diketahui dari penelitian ini yakni *Big 5 Personality* atau Lima Besar Kelompok Karakter menurut (Barrick dan Mount, 1991). Yakni : *Extraversion, Emotional Stability, Agreeableness, Conscientiousness*, dan *Openness to Experience*. Dalam penelitian ini, peneliti mengambil data dari drone empirit. Data diproses untuk mendapatkan klasifikasi 5 besar kelompok dengan tahapan labeling, preprocessing, pembobotan, dan klasifikasi.

Algoritma *Naive Bayes Classifier* dipilih karena memiliki tingkat akurasi yang lebih tinggi. Menurut penelitian yang telah dilakukan oleh (Hamzah, 2014) yang berjudul sentimen analisis untuk memanfaatkan saran kuesioner dalam eva-

luasi pembelajaran dengan menggunakan metode *Naive Bayes Classifier* menyebutkan bahwa metode *Naive Bayes Classifier* dapat memperoleh akurasi mencapai 85,95%. Banyaknya dataset yang dipakai di dalam penelitian ini juga menjadi pertimbangan dalam penggunaan metode *Naive Bayes Classifier*, sebab membutuhkan suatu metode yang mempunyai performansi yang cepat dalam pengklasifikasian serta keakuratan yang cukup tinggi (Larose, D. T., 2006)

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan bahwa permasalahan dalam penelitian ini adalah bagaimana mengklasifikasikan Lima Dimensi Kepribadian masyarakat terhadap pasangan calon Presiden dan calon Wakil Presiden.

1.3. Batasan Masalah

Agar penelitian ini tidak keluar dari pokok pembahasan, maka ruang lingkup penelitian ini dibatasi pada :

1. Data yang digunakan adalah data tweet masyarakat terhadap pasangan calon presiden dan calon wakil presiden yang didapatkan dari sistem drone empirit.
2. Data yang digunakan adalah data yang diunduh pada hari Senin, tanggal 18 Maret 2019, pukul 02:14 PM - 02:29 PM
3. Algoritma klasifikasi yang digunakan adalah *Naive Bayes Classifier*.
4. Metode pembobotan term menggunakan TF dan TF-IDF.
5. Klasifikasi kepribadian sebanyak lima kelas teori kepribadian (*Big Five Per-*

sonality, yaitu Extraversion, Emotional Stability, Agreeableness, Conscientiousness, dan Openness to Experience.

6. Pengambilan data berdasarkan 10 kata kunci : Jokowi, Ma'ruf Amin, Prabowo, Sandiaga Uno, #2019TetapJokowi, #Jokowi2Periode, #JokowiLagi, #2019GantiPresiden, #2019PrabowoPresiden, dan #2019PrabowoSandi.
7. Menggunakan metode *Supervise Learning*.
8. Bahasa pemrograman yang digunakan adalah *Python*.
9. Evaluasi analisis sentimen menggunakan metode *k-fold cross validation*.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka dapat diketahui tujuan dari penelitian ini yakni :

1. Melakukan klasifikasi kepribadian pengguna twitter di Indonesia berdasarkan pendapatnya terhadap pasangan calon Presiden dan Wakil Presiden tahun 2019.

1.5. Manfaat Penelitian

Hasil dari penelitian ini diharapkan membawa manfaat bagi Drone Emprit dan dapat bermanfaat bagi mahasiswa Teknik Informatika untuk dijadikan referensi dalam melakukan penelitian.

1.6. Keaslian Penelitian

Penelitian mengenai klasifikasi sentimen komentar maupun lima besar karakter kepribadian seseorang pada media sosial Twitter, hingga saat ini telah banyak dilakukan oleh peneliti sebelumnya. Namun, berdasarkan referensi dan tinjauan pustaka, penelitian yang diajukan sebagai Tugas Akhir Pendidikan Strata-1 pada program studi Teknik Informatika Fakultas Sains dan Teknologi UIN Sunan Kalijaga Yogyakarta mengenai Implementasi Algoritma *Naive Bayes Classifier* Untuk Menganalisis 5 Besar Karakter Kepribadian Masyarakat Indonesia Terhadap Pasangan Calon Presiden dan Wakil Presiden Tahun 2019 Berdasarkan Data *Tweets*, menggunakan bahasa pemrograman, objek dan *preprocessing* data yang berbeda dan belum pernah dilakukan di UIN Sunan Kalijaga Yogyakarta.

1.7. Sistematika Penulisan

Untuk memberikan gambaran dan kerangka yang jelas mengenai pokok bahasan setiap bab dalam penelitian ini, maka diperlukan sistematika penulisan. Berikut ini gambaran sistematika penulisan pada masing-masing bab :

BAB I : PENDAHULUAN

Bab pertama, merupakan pendahuluan, berisikan penjelasan mengenai latar belakang dilakukannya penelitian, rumusan penelitian, batasan masalah, tujuan penelitian, mafaat penelitian, keaslian penelitian, dan sistematika penulisan penelitian.

BAB II : TINJAUAN PUSTAKA DAN LANDASAN TEORI

Bab kedua, merupakan tinjauan pustaka dan landasan teori, berisikan penjelasan tentang tinjauan pustaka dan landasan teori yang berkaitan dengan topik yang dibahas dalam penelitian ini.

BAB III : METODE PENELITIAN

Bab ketiga, merupakan metode penelitian, berisikan penjelasan mengenai metode ataupun algoritma yang digunakan serta tahapan-tahapan yang dilakukan untuk mencapai tujuan dari penelitian ini.

BAB IV : HASIL DAN PEMBAHASAN

Bab keempat, merupakan hasil dan pembahasan, menjelaskan hasil dari penelitian yang telah dilakukan, dari proses pengambilan data hingga hasil yang dicapai dari penelitian ini.

BAB V : PENUTUP

Bab terakhir, merupakan penutup, berisikan penjelasan tentang kesimpulan yang dihasilkan dari penelitian yang telah dilakukan dan pemberian saran yang dilakukan berdasarkan hasil yang telah dicapai guna dapat menjadi bahan pertimbangan bagi pihak-pihak yang berkepentingan serta untuk kepentingan pengembangan penelitian-penelitian selanjutnya.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan penelitian yang telah dilakukan, didapatkan kesimpulan bahwa Analisis 5 Besar Karakter Masyarakat Indonesia Terhadap Calon Presiden dan Wakil Presiden Tahun 2019 Berdasarkan Komentar Mereka di Media Sosial Twitter dengan algoritma *Naive Bayes Classifier* dapat digunakan untuk mengklasifikasikan data *tweet* dengan tingkat akurasi yang lebih tinggi dengan pembobotan TF yaitu 39,96% dibandingkan dengan metode pembobotan TF-IDF yakni 38,45%. Perhitungan keakurasian ini merupakan hasil rata-rata akurasi dari evaluasi model klasifikasi dengan menggunakan metode *k-fold cross validation*.

Hasil implementasi algoritma *Naive Bayes Classifier* menunjukkan menghasilkan pengelompokan kelas *Openness* sebanyak 42,4%, kelas *Conscientiousness* sebanyak 9%, kelas *Extraversion* sebanyak 14,9%, kelas *Agreeableness* sebanyak 24%, dan kelas *Neuroticism* sebanyak 9,7%.

5.2. Saran

Dari proses penelitian ini masih terdapat banyak kekurangan, oleh karena itu penulis akan memberikan saran yang dapat dijadikan pertimbangan selanjutnya, diantaranya adalah :

1. Peneliti selanjutnya bisa membuat sistem klasifikasi 5 besar karakter seseorang terhadap calon presiden dan wakil presiden pada periode pencoblosan

berikutnya secara otomatis.

2. Diharapkan pada penelitian berikutnya dapat membandingkan dengan metode yang berbeda.
3. Penelitian selanjutnya diharapkan membandingkan akurasi dari beberapa metode klasifikasi yang berbeda.
4. Penelitian selanjutnya diharapkan bisa menggunakan tahapan seleksi data *preprocessing* data yang berbeda dari penulis.

DAFTAR PUSTAKA

- Buntoto, Ghulam Asrofi. 2017. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter.
- Fayyad, Usama. 1996. *Advances in Knowledge Discovery and Data Mining*, MIT Press.
- Fitri, Meisya. 2013. Perancangan Sistem Temu Balik Informasi Dengan Metode Pembobotan Kombinasi Tf-Idf Untuk Pencarian Dokumen Berbahasa Indonesia. Universitas Tanjungpura : Semarang.
- Habibi, Setyohadi, dan Ernawati. 2016. Analisis Sentimen pada Twitter Mahasiswa Menggunakan Metode Backpropagation.
- Han, J., Kamber, M. 2006. *Data Mining: Concepts and Techniques*, University of Illinois at Urbana-Champaign.
- Harsehanto, Ireicca Agustiorini. 2018. Analisis Karakter Kepribadian Seseorang Menggunakan Algoritma *Naive Bayes Classifier* (Studi Kasus Twitter Resmi Basuki Tjahaja Purnama dan Anies Baswedan).
- JUD. 2016. *Pemrograman Python untuk Pemula*. Yogyakarta: CV Jubilee Solusi Enterprise.
- Kusrini, Luthfi, E. T. 2009. *Algoritma Data Mining*. Yogyakarta: ANDI.
- Larose, D. T. 2006. *Nave Bayes Estimation and Bayesian Networks*, in *Data Mining Methods and Models*, John Wiley Sons, Inc., Hoboken, NJ, USA. doi: 10.1002/0471756482.ch5

- Nurhuda, Sihwi, dan Doewes. 2013. Analisis Sentimen Masyarakat terhadap Calon Presiden Indonesia 2014 berdasarkan Opini dari Twitter Menggunakan Metode Naive Bayes Classifier.
- Oktasari, L., Chrisnanto, Y. H., Yuniarti, R. (2016). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Nave Bayes Classifier. Prosiding SNST, 37-42.
- Pan, S., Ni, X., Sun, J., Yang, Q., Chen, Z. 2010. Cross-domain sentiment classification via spectral feature alignment. International World Wide Web Conference Committee.
- Saputro, P. H., Aristian, M., Tyas, D. L. 2017. Klasifikasi Lagu Daerah Indonesia Berdasarkan Lirik Menggunakan Metode Tf- Idf Dan Nave Bayes. Jurnal Teknologi Informatika dan Terapan.
- Sevsa, Bama Apbama. 2018. Analisis Sentimen Komentar pada Indeks Kinerja Dosen Fakultas Sains dan Teknologi UIN Sunan Kalijaga Menggunakan Naive Bayes Classifier.
- Suyanto. 2017. Data Mining untuk Klasifikasi dan Klasterisasi Data. Bandung: Informatika Bandung.